DOCUMENT RESUME

ED 204 378 TM 810 366

AUTHOR Karweit, Nancy L.; Slavin, Robert E.
TITLE Measuring Time-On-Task: Issues of Timing, Sampling and Definition.
INSTITUTION Johns Hopkins Univ., Baltimore, Md. Center for Social Organization of Schools.
SPONS AGENCY National Inst. of Education (ED), Washington, D.C.
REPORT NO CSOS-R-296
PUB DATE Jun 80
GRANT NIE-G-80-0113
NOTE 24p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; *Classroom Observation Techniques; Elementary Education; Elementary School Mathematics; Pretests Posttests; *Research Methodology; *Research Problems; *Sampling; *Time Factors (Learning)
IDENTIFIERS Comprehensive Tests of Basic Skills; *Time on Task

ABSTRACT

How various methodological decisions may influence studies of the effect of time-on-task on achievement are examined. Subjects were students in grades 2-5 in 18 classes taught by 12 teachers in a rural Maryland school district. All students were pre-tested in February 1978 in reading, language arts, math and social studies using the Comprehensive Test of Basic Skills. A post-test was given in May, 1978. It was found that altering definitions of time-on-task to include momentary off-task behaviors affected the conclusions for the importance of time-on-task. Clear evidence was presented that sampling segments of instruction would tend to obscure the positive results for time-on-task. It was also shown that reducing the number of days of observation weakened the effects of time-on-task. However, the timing of the observation was not very important for the noted effects. The effect of sampling fewer than six students was explored and, due to the effect on reliability, it was suggested that this approach would not be advisable. Results suggested that although there is an understandable urge to lessen the observation time in order to bolster the number of settings observed, such steps should only be taken cautiously. (Author/RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Center for Social Organization of Schools

Report No. 296
June 1980

MEASURING TIME-ON-TASK: ISSUES OF TIMING, SAMPLING AND DEFINITION

Nancy L. Karweit and Robert E. Slavin

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. Hollifield

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The Johns Hopkins University

STAFF

Edward L. McDill, Co-Director
James M. McPartland, Co-Director

Karl Alexander
Charles H. Beady
Henry J. Becker
Jomills H. Braddock, II
Ruth H. Carter
Martha A. Cook
Robert L. Crain
Marvin P. Dawkins
Doris R. Entwisle
Joyce L. Epstein
Gail M. Fennessey
James J. Fennessey
Homer D. C. Garcia
Denise C. Gottfredson
Gary D. Gottfredson
Linda S. Gottfredson
Larry J. Griffin
Stephen Hansell
Edward J. Harsch
James A. Harvey
John H. Hollifield
Lawrence F. Howe
Barbara J. Hucksoll
Nancy L. Karweit
Hazel G. Kennedy
Marshall B. Leavey
Nancy A. Madden
Julia B. McClellan
Janice E. McKenzie
Anne McLaren
Phillip R. Morgan
Robert G. Newby
James M. Richards, Jr.
Gail E. Thomas
William T. Trent
Carol A. Weinreich

MEASURING TIME-ON-TASK: ISSUES OF TIMING, SAMPLING AND DEFINITION

Grant No. NIE-G-80-0113

Nancy Karweit

Robert E. Slavin

Report No. 296

June 1980

Published by the Center for Social Organization of Schools, supported in part as a research and development center by funds from the United States National Institute of Education, Department of Health, Education and Welfare. The opinions expressed in this publication do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the Institute should be inferred.

The Johns Hopkins University

Baltimore, Maryland

Introductory Statement

The Center for Social Organization of Schools has two primary objectives; to develop a scientific knowledge of how schools affect their students, and to use this knowledge to develop better school practices and organization.

The Center works through four programs to achieve its objectives. The Studies in School Desegregation program applies the basic theories of social organization of schools to study the internal conditions of desegregated schools, the feasibility of alternative desegregation policies, and the interrelation of school desegregation with other equity issues such as housing and job desegregation. The School Organization program is currently concerned with authority-control structures, task structures, reward systems, and peer group processes in schools. It has produced a large-scale study of the effects of open schools, has developed Student Team Learning Instructional processes for teaching various subjects in elementary and secondary schools, and has produced a computerized system for schoolwide attendance monitoring. The School Process and Career Development program is studying transitions from high school to post secondary institutions and the role of schooling in the development of career plans and the actualization of labor market outcomes. The Studies in Delinquency and School Environments program is examining the interaction of school environments, school experiences, and individual characteristics in relation to in-school and later-life delinquency.

This report, prepared by the School Organization program, examines the methodological problems involved in studies of time-on-task in classrooms.

Abstract

Many recent studies have reported benefical effects of time-on-task on student academic achievement. This paper examines the methodological problems involved in measuring time-on-task, especially problems related to the definition of off-task behavior, length of observation times, days of observations, scheduling of observations, and sampling of students for observation. The findings show that the methodology selected can influence the results of time-on-task studies.

Introduction

Research interest has recently focused on the centrality of time-on-task for understanding classroom effects and effectiveness (Fisher et al., 1976a, 1976b; Filby and Marliave, 1977; McDonald and Elias, 1976; Cooley and Leinhardt, 1978). This research has provided important evidence that links classroom practices, time-on-task and learning outcomes. Although the evidence in general points to positive and meaningful effects of time-on-task, the results are not consistent across studies nor across grade levels/subject matters within studies (e.g. the results obtained in the Beginning Teacher Evaluation Studies, BTES, for mathematics/reading at grades 2/5). Moreover, the effects documented for time-on-task, although positive, have not been uniformly large.[1] Nonetheless, the effects for time factors have assumed appreciable stature by virtue of the fact that time factors can be altered, whereas more statistically important factors, such as family background or entering aptitude, are difficult or even impossible to alter.

Thus, the use of time in classroom continues to be a central theme in educational research. The fact that the results are modest and inconsistent has been attributed to particular methodological or research design problems, not problems with the assumptions guiding the research. That is, the assumption that classroom practices have appreciable impacts on time-on-task which in turn affect the degree of learning is generally not at issue. The present state of encouraging but not entirely clear results is taken to indicate the existence of methodological as opposed to theoretical problems.[2]

Given this slant on the problem, it seems reasonable to ask to what extent the nature of the present findings are due to particular methodological choices or decisions. In particular, it seems useful to explore how

the observation scheme used, the timing of the observation, the length of the observation and the number of observations may affect the detection of time-on-task effects. The present paper uses an existing set of observational data and manipulates it to conform to alternative sampling, timing and definitional choices. Using these alternative choices, we then compare the effects obtained for time-on-task with results reported previously (Karweit and Slavin, 1980). We examine alternative choices in 5 areas:

1. definition of off-task behavior
2. length of observation visit
3. days of observation
4. scheduling of observation
5. sampling of students for observation

Data

The data were collected in four elementary schools in a rural Maryland school district. All schools had open space construction but used essentially traditional methods of classroom organization and instruction. Subjects were students in grades 2-5 in 18 classes taught by 12 teachers. All students were pre- and post-tested in February 1978 in reading, language arts, math and social studies using the Comprehensive Test of Basic Skills. Students in each class were assigned to the top third, middle third, or lower third of the class based on the pre-test information, and two students (one boy and one girl) were chosen from each third for observation. The observations were thus conducted for six students per class, 108 total students, through the second semester of 1978, and the post-test was given in May, 1978.

Students were observed during their mathematics classes, which averaged 50 minutes. Each classroom was observed for at least nine days, and some for

as many as eighteen days. The observers recorded three pieces of information for all six students during a thirty second interval: the nature of the task (procedural, seatwork, or lecture); the student's response to the task (on-task, off-task or no task opportunity), and the content of the instruction (e.g., two digit multiplication, or going over p. 147).

All six students were observed in a predetermined order every thirty seconds. To determine on- or off-task behavior, the observer took a quick look at the student's behavior and recorded the response at that particular instant. The observers were trained not to dwell on deciding whether a behavior was on- or off-task, but to record their first impression in accordance with established definitions of on- and off-task responses.

On average, 100 observations per day were recorded for each student, detailing the task, the content of instruction, and the response. Across all days of observation, we logged about 1000 observation points for each student in the sample, or about 110,000 observation points.

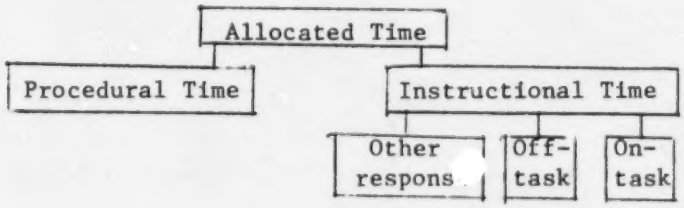
Because of the size of the data base, we entered the task, content and response codes in a summary form which maintained the essentials of the information. Each entry pertained to a specific task or activity and gave the number of seconds each child was on- or off-task during that time. For example, if the class were involved in seatwork during the first ten minutes of the class and then the teacher explained the seatwork during the next eight minutes, we created two entries, one detailing the on/off task behavior during seatwork, and the other giving the on/off task behavior for each of the six children during the teacher-directed activity. From these data, a "day" record was constructed which summarized the daily task, content and

response patterns for each child.

In addition, a special data set containing each 30 second record of task, content and response was compiled for five of the eighteen classrooms. These supplemental data will be used along with the basic data in the analyses.

Definition of On- and Off-Task

On- and off-task behaviors were coded during instructional portions of the lesson only. However, a child could also have a response other than on- or off-task during instructional time. The diagram below depicts the different categories and when they could occur in this observational scheme.



The allocated time was the clock time spent for the mathematics class. Procedural time was any time spent lining up, receiving instructions, being involved in disciplinary action, going to fire drills, being interrupted by the P. A. system and the like. Instruction pertained to the time spent specifically on mathematics instruction; discussion of world events, elections, snow storms and other material not pertaining to math was not coded as instructional time. On-task behavior was defined as behavior appropriate to the task at hand. The definition of appropriate behavior depended upon the task and specific rules of the classroom. "Other response" was used to cover situations in which the child was not on-task but was not off-task

either. Such situations arose when the child was sharpening a pencil, walking to another part of the room to obtain new materials, waiting for the teacher to help with a problem, or doing some other activity because the original assignment was finished.

We focused on two particular problems in assessing off-task behavior. One involved the effect of including momentary off-task behavior; the other involved the effect of including no-task-opportunity time (i.e. "other response") as off-task behavior.

a. Momentary off-task behavior

During any class period, children may momentarily gaze out the window, fidget, or otherwise be momentarily distracted. On the one hand, this momentary off-task time can be looked upon as insignificant for the learning process. On the other hand, momentary off-task behavior may be signalling declining attention and motivation and might therefore be important for understanding the learning process. In the BTES analyses, off-task behaviors shorter than one minute were not counted; i.e. these flickers of inattention were not considered consequential. In coding the data used in the present study, we included all off-task behavior, regardless of duration. To assess whether this decision to include short-term inattention affected the results obtained, we changed all off-task behavior of less than one minute to on-task and repeated the analyses for the supplemental sample of five classrooms. The average rate of on-task behavior increased from .79 to .83 and the standard deviation was reduced from .08 to .07. Including the momentary off-task behavior yielded correlations of .24 between on-task and pre-test score and .45 with post-test score. Excluding momentary off-

task behavior, these correlations became .33 and .39. We carried out regressions of post-test on pre-test and the alternative measures of on-task behavior. Using the measure which excluded momentary off-task behaviors produced more modest results ($p \leq .10$) than did using the more inclusive measure ($p \leq .05$).

Whether momentary inattention is included or not should be based on the particular model of learning one has formulated. Certain views of the learning process may be compatible with inclusion of these momentary distractions; other views would not be. The present exercise was not intended to shed light on whether a particular point of view is proper or improper, but to illustrate that the methodological decision to include/exclude these flickers of inattention affected the results obtained.

b. Other response and off-task behavior

The dichotomy of on- or off-task provides a working categorization of student responses to instruction, but there are numerous ambiguous situations in which the student is not on-task, yet could not be considered off-task. For example, a student may have finished an assignment and have nothing more to do. Students who finish early are likely to be those who need less time, i.e., have more aptitude for the particular task at hand; thus the amount of finished time should be positively related to achievement, in contrast to the negative relationship of off-task variables and achievement. In our data, the correlation between finished time and post-test score was .19 while the correlation between off-task and post-test score was -.28. In regressions (not detailed here) in which finished time was included with off-task time, the effects of off-task time were diminished appreciably.

Length of observation visit

An important design consideration is the length of the observation period. One could observe a single classroom all day long, for some fixed fraction of the day, or for some specific instructional program. Or, a combination of these lengths of observation might be used. Because our interest was in how the use of time affects mathematics achievement, we observed students during their entire mathematics instruction. It was not possible (given our budget constraints on observer time) to observe all teachers within a school. An alternative decision would have been to observe more teachers, but for some smaller segment of their mathematics instruction. We might have decided, for example, that instead of visiting one teacher for sixty minutes we might have used one of the combinations below:

NO. TEACHERS	NO. MINUTES	TOTAL TIME
2	30	60
3	20	60
6	10	60

The choice among these alternatives is basically between getting enough classrooms to provide stable estimates of the effect of time-on-task, and scheduling sufficient time to ensure that the observed behavior is representative. If time-on-task is distributed fairly uniformly across the day or the period of instruction, then a time sample may be entirely adequate. Table 1 gives the means and standard deviations of time-on-task for nine days of observation in one classroom. The first columns provide statistics

---

Table 1 About Here

Table 1

On-Task Rate for Selected Portions
of Mathematics Instruction
in one classroom

Day	Minutes 1 - 10		Minutes 1 - 20		Minutes 1 - 30	
	$\bar{X}$	σ	$\bar{X}$	σ	$\bar{X}$	σ
1	.906	.066	.878	.046	.865	.051
2	.911	.086	.815	.059	.805	.092
3	.922	.078	.923	.079	.921	.094
4	.739	.236	.818	.062	.775	.062
5	.817	.002	.690	.158	.653	.195
6	.889	.087	.869	.073	.804	.099
7	.884	.169	.866	.188	.809	.175
8	.958	.066	.928	.063	.889	.088
9	.825	.196	.841	.148	.850	.171
$\bar{X}$	.872		.848		.819	
$\bar{X}$ this segment			.824		.761	

for the first 10 minutes of class; the second columns for the first 20 minutes, and the third for the first 30 minutes. The overall mean for the time period is supplied as well as the mean for the particular 10 minute segment. The average on-task time in this class was markedly higher during the first 10 minutes of instruction than it was for the next 10 or 20 minutes. Clearly, the timing of observation in this classroom was important for the results obtained as time-on-task was not distributed evenly across the mathematics class time. Other classrooms exhibited different patterns of high and low attention. Some classes started off with lower on-task rates, seemed to warm up to instruction, and have higher on-task rates, and then to die down. Still other classrooms had no consistent pattern at all. Consequently, it is difficult to predict what the effect in general would be if selected portions only of the class time were observed. Thus, although the effect of observing shorter periods may not be consequential for the reliabilities obtained (see Rowley, 1976), how those periods are selected may be very consequential.

To illustrate this point, we regressed post-test achievement scores on pre-test scores and alternate measures of on-task rate, namely measures from the first ten, twenty, thirty and fifty minutes of instruction. The F values obtained for the time-on-task measures were .010, 1.22, 3.09 and 4.34, respectively. The n of this sample was extremely small (22 students): however, the results suggest that observing for shorter segments would have appreciably altered the effects obtained.[3]

Altering the number of days of observation

Conventional wisdom has it that about ten days of observational data should be sufficient to accurately portray the activities of a classroom.

However, few studies have investigated the effects of observing classrooms for fewer or more days, even though this question is of considerable design and practical importance. If we can obtain sufficient information in a shorter period (e.g. five days instead of ten), it would be possible to observe substantially more classrooms without appreciably altering the observation costs.

In the present data set, we observed some classrooms for as many as 18 school days and others for as few as 9 days. With these data, then, we can pretend that we had observed a fixed number of days (e.g. 3, 4, 5, 6, 7, 8, 9) and assess how this observation schedule would have affected the detection of effects of time-on-task on achievement. We think of time-on-task as a variable which is influenced not only by an individual child's disposition, aptitude, and idiosyncracies, but also by the instructional setting in the class and by external events such as the daily weather. Each child may have a stable rate of on-task behavior with daily fluctuations depending upon his response to the classroom and other environmental settings. Given this view of time-on-task as a variable, a natural way to capture the daily and individual variation is to view each day's time-on-task as an item in a scale of total time-on-task. We can then see how consistent the behavior is across a differing number of days or items in the scale.

As expected, increasing the number of days does provide an overall increase in reliability. The median coefficient alphas obtained for 3-9 days were .54, .57, .71, .73, .79, .81, .82. Whether the increase in reliability obtained from observing nine days vs. 5 days is consequential depends on the effect one is trying to document. Because reliability deter-

mines the maximal correlation that one can find between achievement outcomes and time-on-task, the obtained reliability is of some consequence. To assess the effect of these variations in reliability, we used the third grade sample (n=36) and regressed post-test CTBS score on pre-test and alternate measures of time-on-task, namely measures obtained from:

1. five days observation
2. nine days observation
3. eighteen days observation

Table 2 shows that had we observed for the first five or first nine days our effects for time-on-task would have been much more modest. The

---
Table 2 About Here

"18-day" results show significant effects for on-task minutes, engagement rate and off-task rate. Had we observed the same classrooms, but for fewer days, the results obtained would have been much weaker.

It is possible that the days at the end of the observation were significantly different from the days at the beginning; if this were the case we would be witnessing an effect for timing and not for length. This issue is explored in the next section.

Timing of observation days

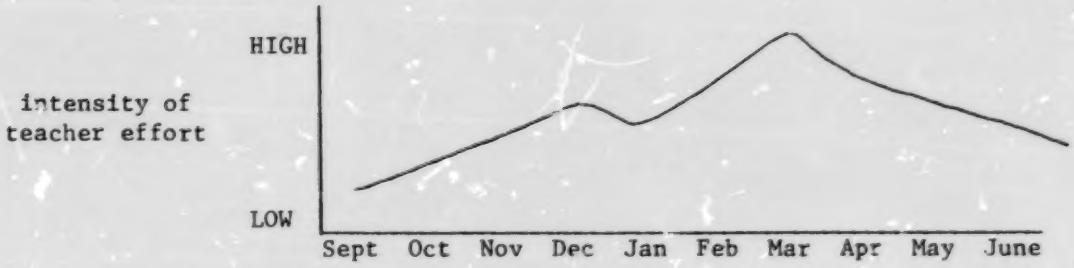
Throughout the school year, there are no doubt more intensive and less intensive times for classroom instruction. For example, one obviously would not want to schedule observations of attending behavior in the few days prior to the Christmas or summer break. Besides these obvious cyclical differences in time-on-task, there may be less obvious sources such as

Table 2

Comparison of Results Obtained
for time-on-task using 5, 9, and 18
days of observation

	18 days		9 days		5 days	
	b/beta	F	b/beta	F	b/beta	F
time-on-task rate	47.51 (.178)	6.56 $p < .05$	33.62 (.129)	3.21 $p < .10$	32.25 (.138)	3.62 $p < .10$
time-on-task minutes	.249 (.165)	5.14 $p < .05$	.156 (.111)	2.28 n.s.	.131 (.110)	2.19 n.s.
time-off-task rate	-48.1 -(.147)	4.33 $p < .05$	-.32.9 -(.10)	2.07 n.s.	-.109 -(.03)	.141 n.s.
time-off-task minutes	-.450 -(.121)	2.86 n.s.	-.329 -(.09)	1.39 n.s.	-16.76 -(.05)	.459 n.s.

different teacher expectations for time-on-task depending upon the time of the year and coverage of material by that point. For example, we might hypothetically view the school calendar from the perspective of the intensity of teacher effort as follows:



Here, January, February, and March are more serious months because of the on-coming deadline of the end of the school year and because there is still time available to redress learning deficiencies.

We are able in a limited fashion to see if time-on-task differs by time of year, using these data. For five classrooms we observed students for a nine-day period in February and also in May. The means and standard deviations for these classrooms are provided in Table 3 for the two different time periods. Table 3 also provides the reliabilities for the two periods of observation (column 5 and 6) and for two mixed scales (S1 and S2) composed

---
Table 3 About Here
---

of equal number of items from February and May. The reliabilities and the means do not appear to be very different for the two time points. This table supplies limited evidence of the consistency of the classroom over time, which suggests that the timing of the observational period may not

Table 3

Comparison of Mean values and reliabilities obtained for time on task in February and May

Feb. Means	May Means	Feb. α	May α	Combined α	S1 α	S2 α
.844	.856	.92	.96	.97	.93	.94
.899	.900	.76	.42	.71	.70	.53
.929	.930	.67	.76	.85	.79	.70
.842	.847	.85	.76	.79	.56	.71

be all that consequential. It also suggests that our failure to find significant effects for time-on-task using only nine days of observational data was most likely due to the decreased reliability of the scale and not due to scheduling effects.

Altering the number of students sampled in the classroom

Another decision which has to be made is whether to observe all students in the room or to follow a sample of students. Whether to observe the entire classroom or selected students depends largely on the purpose of the observation. If one is interested primarily in how classroom organization affects time-on-task, the entire class would probably be observed. Other strictly pragmatic elements such as high absenteeism or sensitivity of identifying students for observation may influence this decision.

Given that the practical and theoretical concerns dictate that sampling should take place, the question is how many students are needed to obtain a reliable estimate of the on-task behavior for the class. We can examine this issue in two ways with these data. In one classroom, we actually observed twelve students as opposed to six, and comparing the class means and standard deviations and reliability obtained for these six vs. twelve shows them to be very similar (x_{12} = .87, x_6 = .86, r_{12} = .92, r_6 = .89). Another way we can focus on this issue is by reducing the number of students and comparing the obtained reliabilities. We used a random selection of three of the six students to assess the effect this sampling might have on reliability. The median reliabilities were not appreciably reduced by selecting only three students.[4] However, given the fragility of time-on-task effects which we have documented here, it would seem worthwhile to keep reliability as high

as possible. In this instance, observing six students would seem desirable.

Summary and discussion

This paper has examined how various methodological decisions may influence studies of the effect of time-on-task on achievement. We found that altering definitions of time-on-task to include momentary off-task behaviors affected the conclusions for the importance of time-on-task. We found clear evidence that sampling segments of instruction would tend to obscure the positive results for time-on-task. We further showed that reducing the number of days of observation also weakened the effects of time-on-task. The timing of the observation was not very important for the noted effects, however. Finally, we briefly explored the effect of sampling fewer than six students and, due to the effect on reliability, suggested that this approach would not be advisable.

The findings in this paper suggest that although there is an understandable urge to lessen the observation time in order to bolster the number of settings observed, such steps should only be taken cautiously. Whether the effects detected and not detected here are bound up with the particulars of this observation study can only be determined by more systematic examination of these methodological issues. In this sense, we hope the paper serves more as a source of what the question might be than of what the answer is. What this paper does show is that methodological decisions, including some that appear quite minor, can have major consequences for the conclusions that are drawn from observational data.

Notes

1. A typical finding has been that time-on-task when added to a regression of post-test on pre-test will increment R^2 by about 3 percent. Although increments to R^2 provide a conservative view of the importance of a variable, other indicators, such as the magnitude of the beta weight or the residual variance accounted for, have not been substantial either.
2. An alternative perspective would be that the work is basically atheoretical so that it is natural to fault the methodology.
3. For five of the eighteen classrooms, we coded each 30 second interval of task, content and response. From this sample, the twenty-two students who had complete test and observational data were used in the regressions reported in this section.
4. The median reliabilities obtained for three students in comparison to six students for three to nine days of observation are:

	3 days	4 days	5 days	6 days	7 days	8 days	9 days
3 Students	.43	.65	.63	.63	.71	.77	.81
6 Students	.54	.57	.71	.73	.79	.81	.82

References

Cooley, W. C. and Leinhardt, G. The Instructional Dimensions Study: The Search for Effective Classroom Processes. Learning Research and Development Corporation, Pittsburgh, 1978.

Filby, N. N. and Marliave, R. Descriptions of Distributions of ACT Within and Across Classes During the A-B Period. (Technical Report IV-1a) Far West Laboratory for Educational Research and Development, 1977.

Fisher, C. W., Filby, N. N., Marliave, R. S., Cohen, L. S., Moore, J. E. and Berliner, D. C. A Study of Instructional Time in Grade 2 Mathematics (Technical Report II-3). San Francisco, Calif.: Far West Laboratory for Educational Research and Development, 1976a.

Fisher, C. W., Filby, N. N., Marliave, R. S., Cohen, L. S., Moore, J. E. and Berliner, D. C. A Study of Instructional Time in Grade 2 Reading (Technical Report II-4). San Francisco, Calif.: Far West Laboratory for Educational Research and Development, 1976b.

Karweit, N. and Slavin, R. Time Factors and Mathematics Achievement. Unpublished paper. February, 1980.

McDonald and Elias, Beginning Teacher Evaluation Study, Phase II 1973-74. Princeton, NJ: Educational Testing Service.

Rowley, G. L. The Reliability of Observational Measures. American Educational Research Journal, 1976, 13, 51-59.